Inbreeding and effective population size of United States Katahdin sheep

G.M. Becker[1], J.M. Burke[2], R.M. Lewis[3], J.E. Miller[4], J.L.M. Morgan[5], B.D. Rosen[6], C.P. Van Tassell[6], D.R. Notter[7], B.M. Murdoch[1*]

[1]Department of Animal, Veterinary and Food Sciences, University of Idaho, 875 Perimeter Dr, 83844, Moscow, ID, United States; [2]USDA, ARS, Dale Bumpers Small Farms Research Center, 6883 AR-23, 72927, Booneville, AR, United States; [3]Department of Animal Science, University of Nebraska–Lincoln, 1400 R St, 68588, Lincoln, NE, United States; [4]Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Skip Bertman Dr, 70803, Baton Rouge, LA, United States; [5]Round Mountain Consulting, 72701, Fayetteville, AR, United States; [6]USDA, ARS, Animal Genomics and Improvement Laboratory, 10300 Baltimore Avenue, 20705, Beltsville, MD, United States; [7]Department of Animal and Poultry Sciences, Virginia Tech, 925 Prices Fork Road, 24060, Blacksburg, VA, United States; *bmurdoch@uidaho.edu (corresponding author)

Abstract

Katahdin sheep are a composite hair breed reared for meat production and known for parasite resistance traits. The aim of this study was to calculate observed heterozygosity (Het_O), inbreeding coefficients and effective population size (Ne) in a study population representing 23 U.S. flocks (N = 4,884 sheep) with 50K genotype data. In the U.S. Katahdin, mean Het_O was 0.37 and mean inbreeding coefficient (Fhat2) was 0.01. Inbreeding coefficient estimates were statistically different (Kruskal-Wallis *P*-value < 2.2e-16) between flocks. The Ne of the study population examined was N = 172 at 13 generations ago, with average linkage disequilibrium r^2 of 0.04 for the generation. Maintaining genetic diversity relies upon maximizing Ne and minimizing inbreeding within a population. The current study found that mean inbreeding of Katahdin sheep was lower than published for other sheep breeds, mean Het_O was greater and Ne was intermediate to estimates for other breeds.

Introduction

Genetic diversity is important for the preservation of livestock species. Inbreeding can result in increased homozygosity, reduction of genetic diversity and the accumulation of detrimental alleles (Vostry et al., 2018). Negative outcomes related to increased homozygosity can also occur through inbreeding depression, which can affect economically important traits such as birth weight, weaning weight, average daily gain and litter size (Gholizadeh and Ghafouri-Kesbi, 2016; Selvaggi et al., 2010). The Katahdin breed of sheep was developed in the later half of the twentieth century (Wildeus 1997) and has since grown to become economically important to the United States sheep industry (Thorne et al., 2021). The aim of this study was to characterize observed heterozygosity (Het_O), inbreeding and effective population (Ne) size in a large population of 4,884 Katahdin sheep with 50K genotype data.

Materials & Methods

Twenty-three U.S. Katahdin flocks enrolled with the National Sheep Improvement Program (NSIP) participated in the current study. Sheep producers collected blood cards and DNA samples were extracted and genotyped with the Neogen GGP Ovine 50K genotype array. All autosomal markers were filtered by the following exclusionary parameters: sample call rate <

0.90, marker call rate < 0.90, marker minor allele frequency < 0.01. Sex chromosome markers were not included in analyses. In total, 37,856 markers were used for principal component analysis (PCA) and effective population (Ne) size estimation. Data for heterozygosity and inbreeding analyses were pruned to remove markers in high linkage-disequilibrium (LD) using the --indep-pairwise function in plinkv1.9 with the parameters window size of 50 kb, variant count moving five markers per window and r^2 threshold of 0.8. Following LD-pruning, 34,772 autosomal markers were used in inbreeding and mean heterozygosity analyses. Observed heterozygosity, inbreeding and PCA analyses were conducted using plink v1.9 with the functions --het, --ibc and –pca, respectively (Chang et al., 2015). Ne analysis using LD-based methods were conducted with SNeP1.11 with default parameters (Barbato et al., 2015). R version 4.1.1 was used for data visualization with packages 'ggpubr' and 'ggplot2' (Kassambara 2020; Wickham 2016). Statistical significance of (Fhat2 ~ Flock) was evaluated with the Kruskal-Wallis test. The r package 'rcompanion' was used to calculate 95% confidence intervals for inbreeding coefficients (Mangiafico 2020).

Results

The structure of the study population was investigated by PCA (Figure 1). The flocks in this study were located across the five geographical regions of the U.S.: the East Coast (EC), Gulf Coast (GC), Midwest (MW) and the West Coast/Rocky Mountain (WR) regions. There was greater diversity of sheep from the GC region in comparison to sheep from other regions.

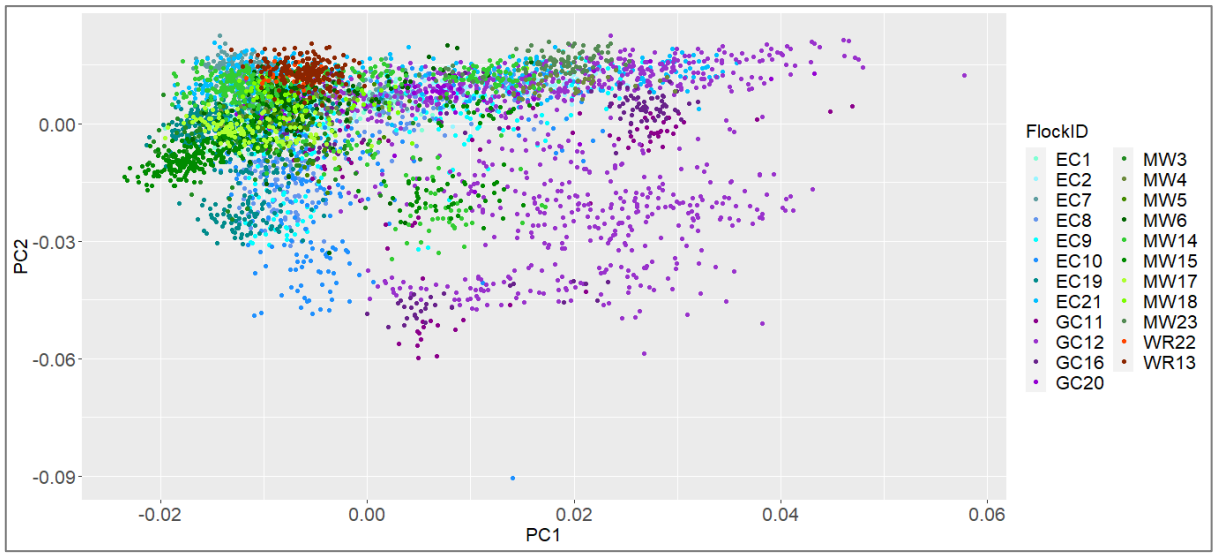

Figure 1. Principal component analysis (PCA) for U.S. Katahdin sheep. Animals are color-coded by region (East Coast in blue shades, n = 2,167; Gulf Coast in purple shades, n = 1,088; Midwest in green shades, n = 1,414; West Coast/Rocky Mountain in orange shades, n = 215) and labelled by flock from 1 to 23, with a total of 4,884 sheep.

The mean heterozygosity observed (Het_O) in the study population was 37%. This was found to be higher than mean Het_O for previously reported breeds (Table 1). The range in mean Het_O by flock was 0.36 to 0.39 with a standard deviation of 0.01. The range in Het_O for individual sheep of all flocks was 0.25 to 0.45 with a standard deviation of 0.02.

Population and flock inbreeding coefficients were calculated using Fhat2, which is calculated (observed homozygous – expected) / (1 – expected) (Rovelli et al., 2021). Individual sheep inbreeding coefficients for all flocks ranged from -0.42 to 0.34 with a standard deviation of

0.07, and the mean inbreeding coefficients by flock ranged from -0.1 to 0.07 with a standard deviation 0.04 (Figure 2). Inbreeding coefficients differed significantly across flocks (Kruskal-Wallis *P*-value < 2.2e-16). The mean inbreeding coefficient over the entire study population was 0.01 and all inbreeding coefficient mean calculations were supported by 95% confidence intervals.

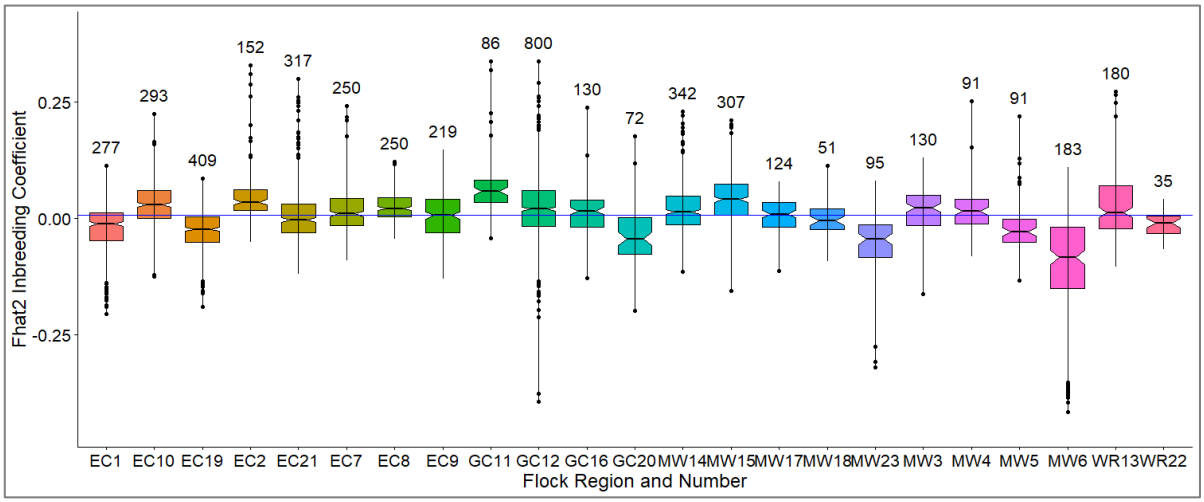

Figure 2. Fhat2 inbreeding coefficients by Katahdin flock. Overall population mean is given by the blue line and individual flock size (n) are given above each boxplot.

Fhat2 inbreeding coefficients have been previously estimated in other sheep breeds (Davenport et al., 2020; Rochus et al., 2020). The mean inbreeding coefficient calculated for this study population was smaller than mean estimates published for other breed populations (Table 1). Inbreeding coefficients may be difficult to directly compare across studies due to differences in sample structure and size.

Table 1. Inbreeding coefficient and observed heterozygosity compared to other published estimates.

		Rochus et al., 2020					Davenport et al., 2020					
	KT*	D	F	G	Gute	K	H	S	WS	O	SH	R
Number of Animals	4,884	21	10	19	22	21	45	68	37	11	44	43
Inbreeding Coefficient	0.01	0.3	0.36	0.03	0.25	0.26	0.14	0.13	0.14	0.05	0.09	0.16
Mean Observed Heterozygosity	0.37	0.30	0.30	0.33	0.31	0.34	0.33	0.33	0.34	0.35	0.34	0.30

*Data from this study
KT-Katahdin; F-Fjällnäs; G-Gotland; K-Klövsjö; H-Hampshire; S-Suffolk; WS-Western Suffolk; O-Oxford; SH-Shropshire; R-Rambouillet.

For this population, historic Ne was estimated to be 901 (as of 234 generations ago) while the most recent Ne estimate was 172 (as of 13 generations ago). The average LD per bin decreased from r^2 of 0.12 (standard deviation 0.16) at 234 generations ago to r^2 of 0.04 (standard deviation 0.06) at 13 generations ago.

Discussion

This study encompasses nearly 5,000 sheep from 23 Katahdin flocks located across the United States. Some of the sheep in this study exhibited inbreeding coefficients of 6.25% or greater, indicating cousin mating (Li et al., 2011); however, the mean inbreeding coefficient was smaller than estimates for other breeds and the mean observed heterozygosity was higher. Previous estimations of Ne for other breeds range from 25 to 1,317 (Kijas et al., 2012; Pasandideh et al., 2020). Two of the founding breeds of the Katahdin, the Wiltshire and Suffolk, have previously reported Ne estimates of 100 and 300/569 (Irish/Australian Suffolk), respectively (Kijas et al., 2012). In the current study, 4,884 Katahdin sheep were evaluated and found to have an estimated Ne that falls intermediate of other breed reports. These statistics have allowed for characterization of genetic diversity within a robust sample population of U.S. Katahdin sheep.

References

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015) Front. Genet. 6:109. https://doi.org/10.3389/fgene.2015.00109

Chang, CC., Chow, CC., Tellier, LCAM., Vattikuti, S., et al. (2015) GigaScience 4(1). https://doi.org/10.1186/s13742-015-0047-8

Davenport, K. M., Hiemke, C., McKay, S. D., Thorne, J. W., et al. (2020) An. Genet. 51(2): 284-291. https://doi.org/10.1111/age.12905

Gholizadeh, M., & Ghafouri-Kesbi, F. (2016) Sm. Rum. Research 144: 184-190. https://doi.org/10.1016/j.smallrumres.2016.09.012

Kassambara, Alboukadel. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. Available at: https://CRAN.R-project.org/package=ggpubr

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., et al. & International Sheep Genomics Consortium. (2012) PLoS Bio., 10(2). https://doi.org/10.1371/journal.pbio.1001258

Li, M. H., Strandén, I., Tiirikka, T., Sevón-Aimonen, M. L., et al. (2011) PLoS One, 6(11): e26256. https://doi.org/10.1371/journal.pone.0026256

Mangiafico, Salvatore. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.4.13. Available at: https://CRAN.R-project.org/package=rcompanion

Pasandideh, M., Gholizadeh, M., & Rahimi Mianji, G. (2020). Iranian J. App. An. Sci., 10(2): 287-295.

Rochus, C. M., Jonas, E., & Johansson, A. M. (2020) BMC Genet. 21(1): 1-9. https://doi.org/10.1186/s12863-020-0827-8

Rovelli, G., Luigi-Sierra, M. G., Guan, D., Sbarra, F., et al. (2021) PeerJ, 9:e12049. https://doi.org/10.7717/peerj.12049

elvaggi, M., Dario, C., Peretti, V., Ciotola, F., et al. (2010) Sm. Rum. Research, 89(1): 42-46. https://doi.org/10.1016/j.smallrumres.2009.12.005

Thorne, J. W., Murdoch, B. M., Freking, B. A., Redden, R. R., et al. (2021) An. Genet., 52(4): 395-408. https://doi.org/10.1111/age.13067

Vostry, L., Milerski, M., Schmidova, J., & Vostra-Vydrova, H. (2018) Sm. Rum. Research 168: 25-31. https://doi.org/10.1016/j.smallrumres.2018.09.004

Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, Available at: https://ggplot2.tidyverse.org.

Wildeus, S. (1997) Jour. An. Sci., 75(3): 630-640. https://doi.org/10.2527/1997.753630x